

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-044560

(43)Date of publication of application : 14.02.1995

(51)Int.Cl.

G06F 17/27

(21)Application number : 05-190998

(71)Applicant : HITACHI LTD  
HITACHI SOFTWARE ENG CO LTD

(22)Date of filing : 02.08.1993

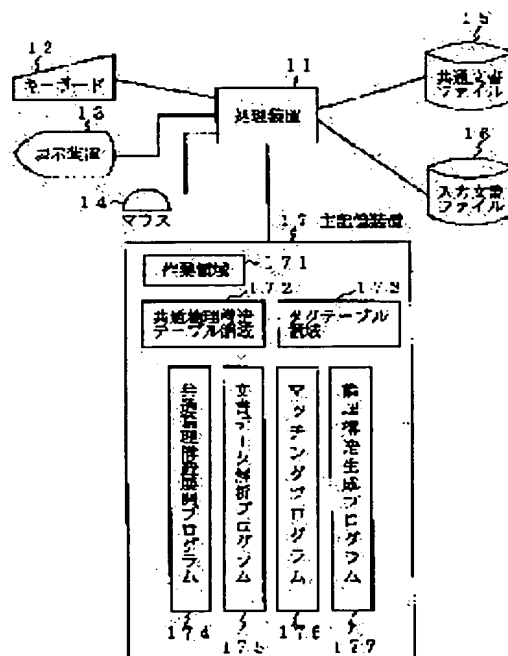
(72)Inventor : SUMIZAWA TAE  
YAMAZAKI EIJI

## (54) LOGICAL STRUCTURE RECOGNITION PROCESSING SYSTEM IN DOCUMENT PROCESSOR

## (57)Abstract:

PURPOSE: To efficiently prepare a routine document corresponding to a fixed rule.

CONSTITUTION: A common document file 15 stores the appearance order rule of the logical structure of a document and an input document file 16 stores the document prepared based on a general purpose markup language rule. A program 174 develops the rule of the file 15 in a chain form common logical structure table and stores it in an area 172 and the program 175 acquires document tags from markup documents fetched to an operation area 171 in order and sets them in the area 173 as a tag table. The program 176 refers to the logical structure table, judges whether or not the appearance order of the document tags of the tag table matches with the rule, extends the chain of the tag table when it matches with the rule and points out the effect to a user when it does not match with the rule. After the matching of the entire document tags is judged, the program 177 generates the logical structure of the document.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-44560

(43) 公開日 平成7年(1995)2月14日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/27		7315-5L	G 0 6 F 15/ 20	5 5 0 E

審査請求 未請求 請求項の数 3 O L (全 10 頁)

(21) 出願番号 特願平5-190998

(22) 出願日 平成5年(1993)8月2日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(71) 出願人 000233055

日立ソフトウェアエンジニアリング株式会社

神奈川県横浜市中区尾上町6丁目81番地

(72) 発明者 住澤 妙

神奈川県横浜市中区尾上町6丁目81番地

日立ソフトウェアエンジニアリング株式会社内

(74) 代理人 弁理士 鈴木 誠

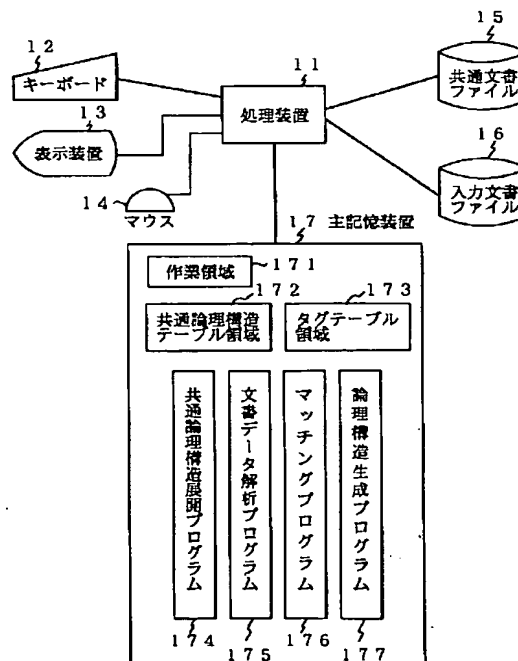
最終頁に続く

(54) 【発明の名称】 文書処理装置における論理構造認識処理方式

(57) 【要約】

【目的】 一定のルールに従った定型文書を効率よく作成する。

【構成】 共通文書ファイル15は、文書の論理的構造の出現順序ルールを記憶する。入力文書ファイル16は汎用マークアップ言語規約に基づいて作成された文書を記憶する。プログラム174はファイル15のルールをチェーン形式の共通論理構造テーブルに展開して領域172に記憶する。プログラム175は作業領域171に取り込まれたマークアップ文書から文書タグを順番に取得し、タグテーブルとして領域173に設定する。プログラム176は論理構造テーブルを参照して、タグテーブルの文書タグの出現順序がルールに合致するか判定し、合致すればタグテーブルのチェーンをのぼし、合致しなければその旨を利用者に指摘する。全文書タグの合致判定後、プログラム177は文書の論理構造を生成する。



## 【特許請求の範囲】

【請求項1】 論理要素の名称を表わすタグ名称である文字列とその前後のタグ区切り記号とからなる文書タグ及びそれに続くテキスト文によって表わされるマークアップランゲージに従って作成された文書（以下、マークアップ文書と称す）を処理する文書処理装置における論理構造認識処理方式であって、

作成されたマークアップ文書を記憶する手段と、  
予め定義された文書の論理的構造の出現順序に関するルールを記憶する手段と、

前記の論理的構造の出現順序に関するルールをチェーン形式の共通論理構造テーブルに展開する手段と、

前記マークアップ文書から文書タグを順次取得する手段と、

前記共通論理構造テーブルを参照して、前記文書タグの出現順序が前記ルールと合致するか判定し、合致しないときはその旨を出力する手段と、を有することを特徴とする論理構造認識処理方式。

【請求項2】 請求項1記載の文書処理装置における論理構造認識処理方式において、文書タグの出現順序の判定に先立って、そのタグ名称の適否を判定することとを特徴とする論理構造認識処理方式。

【請求項3】 請求項1および2記載の文書処理装置における論理構造認識処理方式において、文書タグのタグ名称、出現順序が正しい場合、当該マークアップ文書に固有の特定論理構造を生成することを特徴とする論理構造認識処理方式。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、文書の作成・編集を行う文書処理装置における論理構造認識処理方式に係り、特に、文書をマーク付けする汎用マークアップ言語規約に従った文書タグを用いて作成された文書データの論理構造認識処理方式に関する。

## 【0002】

【従来の技術】文書の内容は、章・節・段落等の文書の持つ意味について着目した論理構造と、ページやフレーム等の文書の内容を物理的に配置するための割付け構造の二つに大別することができる。

【0003】従来の技術は、この点に着目し、構造文書の作成を行っている。例えば、特開平3-127170号公報には、論理構造とその内容を並列して作成し、文書に固有の論理的構造を作成することが記載されている。この従来技術では、操作者にガイダンスを示すことで次内容を促し、操作性の向上をはかっている。

## 【0004】

【発明が解決しようとする課題】ワードプロセッサの普及に伴い文書処理システムは急速に進歩し、また、ワークステーションを中心に国際標準・オープン化への対応が強く必要とされている。これらの文書表現は、ISO

で規格化された交換形式表現や、同じくISOで規格化されたマークアップゲージに代表される。このうち、ISOの交換形式は、実装面での負荷が大きい等の点から、文字コードだけで表現できるマークアップゲージによる文書表現の普及が予想され、この表現方法で一定のルールに従った文書を容易に作成できることが要求されてきた。

【0005】しかしながら、従来技術は、先の特開平3-127170号公報にも記載されているように、操作者がガイダンスに従って文書内容を入力すると、それに伴い文書の論理的構造を順次生成していくもので、処理装置は文書の論理的構造単位の内容をキー入力によって取得する方式であり、一定のルールに従った定型文書を効率よく作成するには不向きである。

【0006】本発明の目的は、文書処理装置において、汎用マークアップ言語規約に従って作成された入力文書データから、その文書の固有の論理的構造を容易に生成することにある。

## 【0007】

【課題を解決するための手段】上記目的を達成するために、請求項1の発明は、論理要素の名称を表わすタグ名称である文字列とその前後のタグ区切り記号とからなる文書タグ及びそれに続くテキスト文によって表わされるマークアップランゲージに従って作成された文書（マークアップ文書）を処理する文書処理装置において、作成されたマークアップ文書を記憶する手段と、予め定義された文書の論理的構造の出現順序に関するルールを記憶する手段と、前記論理的構造の出現順序に関するルールをチェーン形式の共通論理構造テーブルに展開する手段と、前記マークアップ文書から文書タグを順次取得する手段と、前記共通論理構造テーブルを参照して、前記文書タグの出現順序が前記ルールと合致するか判定し、合致しないときはその旨を出力する手段とを有することを特徴とする。

【0008】請求項2の発明は、文書タグの出現順序の判定に先立って、そのタグ名称の適否を判定することとを特徴とする。

【0009】請求項3の発明は、文書タグのタグ名称、出現順序が正しい場合、当該マークアップ文書に固有の特定論理構造を生成することを特徴とする。

## 【0010】

【作用】必要とされる予め定義された論理的構造の出現順序に関するルールの情報を共通論理構造テーブルとして、内部的にランダムアクセスメモリ等に展開する。これにより、マークアップされた入力文書データの文書タグの出現順序と共通論理構造テーブルのマッチングをとることで、容易に文書タグの出現順序がルールと合致するかどうか判定でき、処理の迅速化が図れる。また、この時、共通論理構造テーブルのルールを表わす情報の中に含まれている文書タグに対応する名称を利用して、入

力文書データの文書タグの出現順序の判定に先立って、そのタグ名称の適否を判定することもできる。

【0011】マークアップされた入力文書データの文書タグ名称の出現順序等と共通論理構造テーブルのマッチングをとり、適合しない文書タグが現れた場合にはそれを却下して、利用者に対し誤りである旨を指摘する。これに応じ利用者は指摘された誤りを修正する。このようにして、入力文書データの文書タグの出現順序等が一定のルールに従っていることが判定された場合、該入力文書に固有の論理構造を生成する。

【0012】

【実施例】以下、本発明の一実施例を図面を用いて詳述する。

【0013】図1は本発明の一実施例の構成図を示したものである。図中、11は種々の処理を実行する処理装置(CPU)、12は処理実行コマンド等を入力する為に使用するキーボード、13は出力結果を表示する為の表示装置、14は表示画面上での入力指示に使用するマウス、15は予め定義された文書の論理的構造の出現順序に関するルール(例えば、ISOの国際規格であるODA/事務文書体系規約に準拠するルール)を前もって記憶しておく共通文書ファイル、16は汎用マークアップ言語規約(例えば、文書構造の記述方法に関する国際規格であるSGML/汎用マークアップ言語規約)に基づいて作成された文書を記憶する入力文書ファイル、17は処理装置11が実行するプログラムやデータを保持するランダムアクセスメモリの主記憶装置である。

【0014】主記憶装置17には、データ領域として作業領域171、共通論理構造テーブル領域172およびタグテーブル領域173があり、プログラムとしては共通論理構造展開プログラム174、文書データ解析プログラム175、マッチングプログラム176および論理構造生成プログラム177がある。プログラム174は、共通文書ファイル15の予め定義された論理的構造の出現順序に関するルールをチェーン形式の共通論理構造テーブルとして領域172に展開するプログラム、プログラム175は、入力文書ファイル16の汎用マークアップ言語規約に従って作成された入力文書データから、文書タグ情報を識別し、タグテーブルとして領域173に展開するプログラム、プログラム176は、プログラム175によって生成されたタグテーブルのパスを元に、プログラム174によって展開された共通論理構造テーブルを検索し、論理構造の下部構造をルールに従って評価・決定することによって、マークアップされた入力文書データのタグ名称の出現順序と共通論理構造のマッチングを図り、併せて文書タグが論理構造上、正確に設定されているかどうか検査するプログラム、プログラム176は入力文書ファイル16中の入力文書に固有の文書の論理的構造を生成するプログラムである。

【0015】初めに、図2のフローチャートを用いて本

発明による論理構造認識処理の全体的流れを説明する。まず、入力文書ファイル16及び共通文書ファイル15をオープンする(ステップ201、202)。次に、プログラム174を起動し、共通文書ファイル15中の必要とされる予め定義された論理的構造の出現順序に関するルールの情報を共通論理構造テーブルとして、主記憶装置17の共通論理構造テーブル領域172に展開する(ステップ203)。次に、入力文書ファイル16の汎用マークアップ言語規約に従って作成された入力文書データを読み込んで主記憶装置17の作業領域171に格納する(ステップ204)。この入力文書データの読み込みが終了すると(ステップ205)、プログラム175、176を起動する。プログラム175は作業領域171の入力文書データを解析し、文書タグ情報を抽出し、タグテーブルとしてタグテーブル領域173に設定する(ステップ206)。プログラム176は、このタグテーブル領域173を参照し、タグの出現順序を上位構造から下位構造までを1つのパスとして、該タグテーブルに登録する。さらに、プログラム176は、領域172の共通論理構造テーブルを検索して、論理構造の下部構造をルールに従って評価・決定し、タグテーブルに登録した順序で下位構造が作成できるかチェックすることによって文書タグと共通論理構造のマッチングを図る(ステップ207)。この時、設定された文書タグが予め定義された論理的構造の出現順序に関するルールに則していない場合、その旨を表示装置13に出力する。最後の文書タグ名までの処理が終了すると(ステップ208)、プログラム177を起動する。プログラム177は領域173のタグテーブルに従い、入力文書の論理的構造を生成する(ステップ209)。最後に、共通文書ファイル15および入力文書ファイル16をクローズにする(ステップ210、211)。

【0016】図3は文書構造の出現順序に関するルールを示す共通論理構造の例である。共通論理構造には、各ノードに固有の名称を利用者可視名として付けておき(例えば“文書”、“和文表題”など)、これをマークアップに利用し、マークアップ文書の文書タグと該利用者可視名を一件一件対応させ、論理的構造の出現順序に関するルールをこの共通論理構造に記述されたものと対応するかチェックする。

【0017】図4は、図3の共通論理構造から利用者可視名、論理構造の出現順序に関するルール及び、識別子等の情報を取得して、主記憶装置17の領域172に共通論理構造テーブルとしてチェーン形式に展開したものである。図3から図4への変換は、共通論理構造展開プログラム174によって行われる。このように、共通論理構造の情報を共通文書ファイル15から内部のランダムアクセスメモリ等の領域172へ展開することによって、共通論理構造の情報を取得する際にアクセスが速くなるため、処理速度の向上が望める上に、共通論理構造

から入力文書の特定期理構造を生成する際に必要となる  
 論理構造の上位と下位構造のパスの接続を示す識別子の  
 リストを取得することができる。図4中のSEQ・REP・OPTは、それぞれ下位構造が順番に現れること・  
 繰り返すこと・省略可能となることを表している。例え  
 ば、識別子“本体”配下には「SEQ 3 4 5 6  
 (OPT 7)」のルールに基づいてノードが生成される  
 ことになっており、更に識別子3“表題”の配下には  
 「SEQ 8 9」の順で最下位オブジェクトである識  
 別子8“和文表題”、識別子9“英文表題”がそれぞれ

10 順に生成されるというルールが記述されている。  
 【0018】図5は、マークアップされた入力文書デー  
 タから文書タグの情報を抜き出し文書タグテーブルとし  
 て、文書タグ名その他の必要とされる情報を設定したテ  
 ーブル・チェーンを示したものである。

【0019】文書タグは、図5の(b)に拡大して示す  
 ように、論理要素の名称を表すマークであることを示す  
 為に用いられるタグ区切り記号と、それに付随する論理  
 要素の名称を表すタグ名称である文字列及び、それに続  
 く文書の内容であるテキストとの論理要素の区切りを表  
 すタグ閉じ記号からなる。

【0020】文書データ解析プログラム172では、作  
 業領域171に取り込まれた図5の(a)に示すような  
 入力マークアップ文書データから、データと論理構造単  
 位の名称との区切りを示す記号と、論理構造の名称と、  
 続くデータとの区切りを示す記号を識別することで、論  
 理構造単位の名称である文書タグ名を取得し、タグテー  
 ブル領域173に設定を行う。この際、共通論理構造の  
 すべてのノードに対してタグを対応させると、必然的に  
 決められてしまうタグも入力しなくてはならず煩雑であ  
 る。そこで、汎用マークアップ言語規約では次の様な構  
 造的ルールを設けている。

① 利用者可視名でノードとルートからのノード列が一  
 意に決まるもので、上位の文書タグを省略できる。

② 下位構造が選択の余地なく一意に決まる文書タグを  
 省略できる。

このように設定することで、利用者は共通論理構造で規  
 定されている全てのノードを文書タグとして記す必要が  
 なくなり、煩雑さから開放されることになる。図5の  
 (c)に示すタグテーブルで、斜線を施さないノード

40 が省略可を示している。  
 【0021】次のマッチングプログラム176では、省  
 略されたノードを補正(追加)し、文書タグの出現順序  
 等が予め定義されたルールである共通論理構造に反する  
 ものである場合はエラーを出力し、適合する場合はタグ  
 テーブルのチェーンを繋げ、共通論理構造に沿った形で  
 文書タグテーブルのパスを生成する処理を行う。

【0022】図6にマッチング処理のフローチャートを  
 示す。以下では、図7の例を用いて共通論理構造と文書  
 タグテーブルのチェーンのマッチング処理の概要を説明

す。

【0023】いま、“文書”“和文表題”“英文表題”  
 “和文著者名”と順にタグテーブルのチェーンが生成さ  
 れ、次に“和文勤務先”の文書タグが文書データ解析プ  
 ログラム175によって取得されたとする。マッチング  
 プログラム176は共通論理構造テーブルのカレントポ  
 インタをノード2に設定した後(ステップ601)、該  
 ノード2以降をサーチし(ステップ602、604、6  
 05)、処理対象文書タグ“和文勤務先”のノード識別  
 子12を取得する(ステップ606)。この時、綴り上  
 の誤り等により、求める文書タグ名と一致する利用者可  
 視名が共通論理構造テーブル上に存在しない時は、その  
 旨をエラーメッセージとして出力する(ステップ60  
 3)。これにより、利用者は指摘された誤りを修正すべ  
 ばよい。

【0024】文書タグが正しく取得された場合には、そ  
 の時のタグテーブルの最終位置を退避した後(ステップ  
 607)、タグテーブル及び共通論理構造テーブルの各  
 カレントポインタを戻して(ステップ608、60  
 9)、ステップ607で退避したタグテーブルの最終位  
 置まで再び順に共通論理構造テーブル及びタグテー  
 ブルを検索し(ステップ610、611、613、61  
 4)、文書タグの生成過程が正しく、処理対象文書タグ  
 に対応するタグテーブルを続けて生成することができ  
 るか判定を行う。図7では、カレントポインタのある共通  
 論理構造テーブルのノード2配下に記述された下位構造  
 の出現順序に関するルールは「SEQ 3 4 5 6  
 (OPT 7)」であるので、この順に各テーブルを検索  
 し、タグテーブルが出現順序に関するルールに従って生  
 成されることを確認する。すると、ノード識別子5の  
 “勤務先”の配下に、ノード識別子12“和文勤務先”  
 が発見され、それまでの従属関係が正しいものであるこ  
 とが判定される。

【0025】文書タグの生成過程が正しい場合、省略さ  
 れているノードを追加等するため、それまで生成されて  
 いたタグテーブルを補正する必要があるか判定し(ステ  
 ップ615)、補正要の場合はタグテーブルのチェーン  
 を補正後(ステップ616)、取得した文書タグを接続  
 する(ステップ617)。図7の場合、取得した“和文  
 勤務先”の文書タグは、“勤務先”のノードを追加し  
 て、途中まで生成されていたタグテーブルのチェーンを  
 補正後、接続されることになる。

【0026】一方、上述の文書タグの続きとして“英文  
 著者名”の後に“和文勤務先”ではなく、“英文勤務  
 先”の文書タグが出現した場合、ノード識別子5の“勤  
 務先”の配下には、シーケンスとしてノード識別子12  
 “和文勤務先”、続いてノード識別子13“英文勤務  
 先”が来るべきであることから、ステップ611におい  
 て“和文勤務先”のタグより前に出現する“英文勤務  
 先”の文書タグの誤りが指摘され、マッチングプログラ

ム176は、その旨をエラーメッセージとして表示する(ステップ612)。これを見て利用者は誤りを修正する。

【0027】このように、文書データ解析プログラム175が文書タグを取得する度にマッチングプログラム176が動作し、文書タグの構造上の適否を判定した上で、文書タグの出現順序等が正しければ、タグテーブルに該文書タグデータをチェーンする処理が行われる。

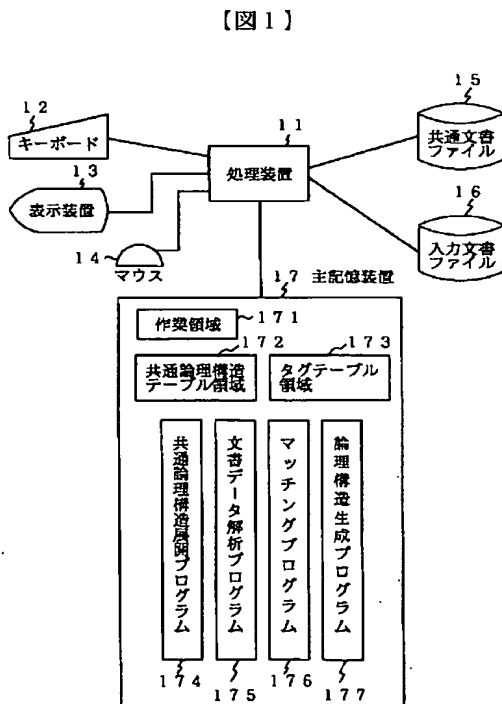
【0028】論理構造生成プログラム177は、全ての文書タグの出現順序が一定のルールに適合していることが判定された場合、プログラム176からの要求により、タグテーブル等をもとに入力文書の特定論理構造を生成する。これは、基本的には図4から図3を生成する処理であり、タグテーブルが正しくチェーンされていれば、それを元に入力文書に固有の特定論理構造を生成することはたやすい。

【0029】

【発明の効果】以上の説明から明らかであるように、本発明によればマークアップページに従って作成された文書を入力した場合、その文書タグの出現順序等が一定のルールに合致しているかどうか迅速に認識し、誤ったマークアップがあればそれを利用者に指摘することにより、一定のルールに従った定型文書を効率よく作成することができる。

【図面の簡単な説明】

【図1】本発明による文書処理装置の一実施例の構成図＊



＊である。

【図2】本発明による論理構造認識処理の流れを説明するためのフローチャートの一例である。

【図3】共通論理構造の一例を示す図である。

【図4】共通論理構造をチェーン形式に展開した図である。

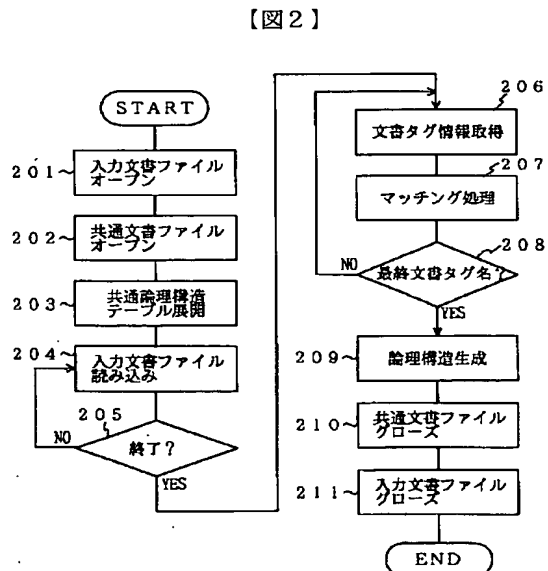
【図5】文書データ解析処理の概要を示す図である。

【図6】マッチング処理の流れを説明するためのフローチャートの一例である。

10 【図7】マッチング処理の概要を示す図である。

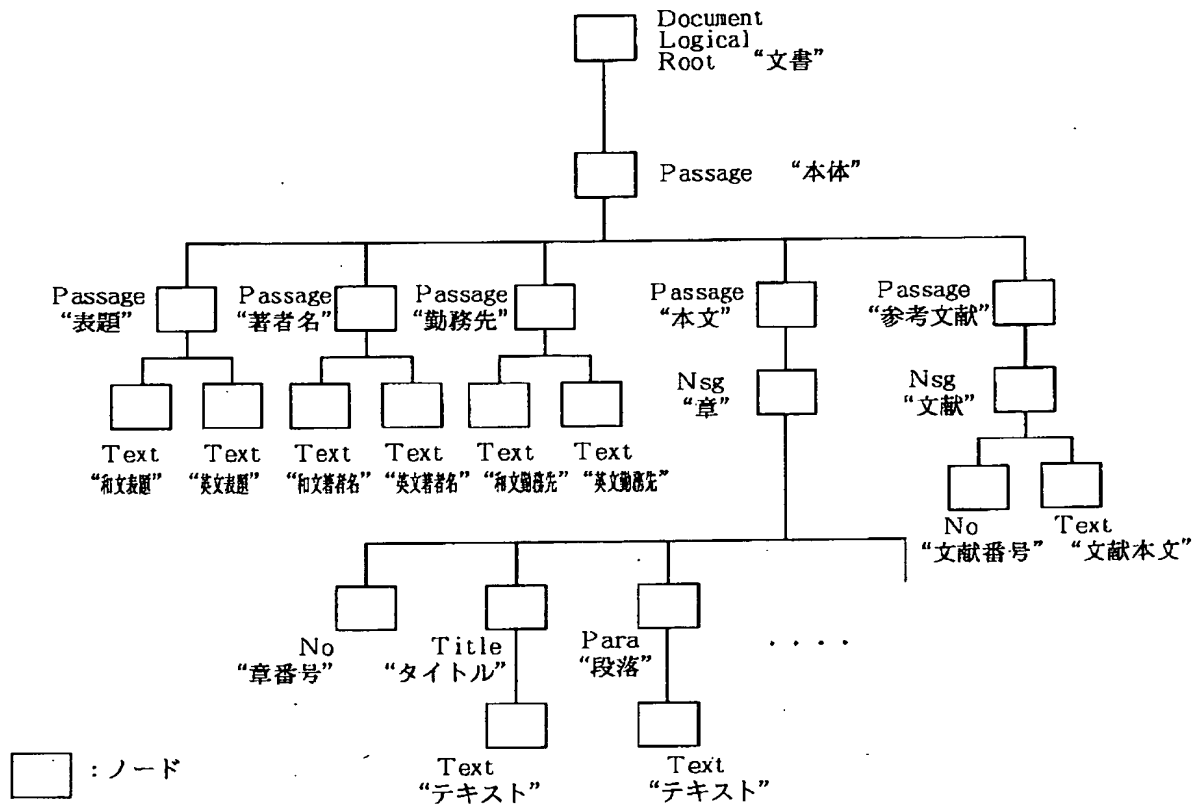
【符号の説明】

- 11 処理装置
- 12 キーボード
- 13 表示装置
- 14 マウス
- 15 共通文書ファイル
- 16 入力文書ファイル
- 17 主記憶装置
- 171 作業領域
- 172 共通論理構造テーブル領域
- 173 タグテーブル領域
- 174 共通論理構造展開プログラム
- 175 文書データ解析プログラム
- 176 マッチングプログラム
- 177 論理構造生成プログラム



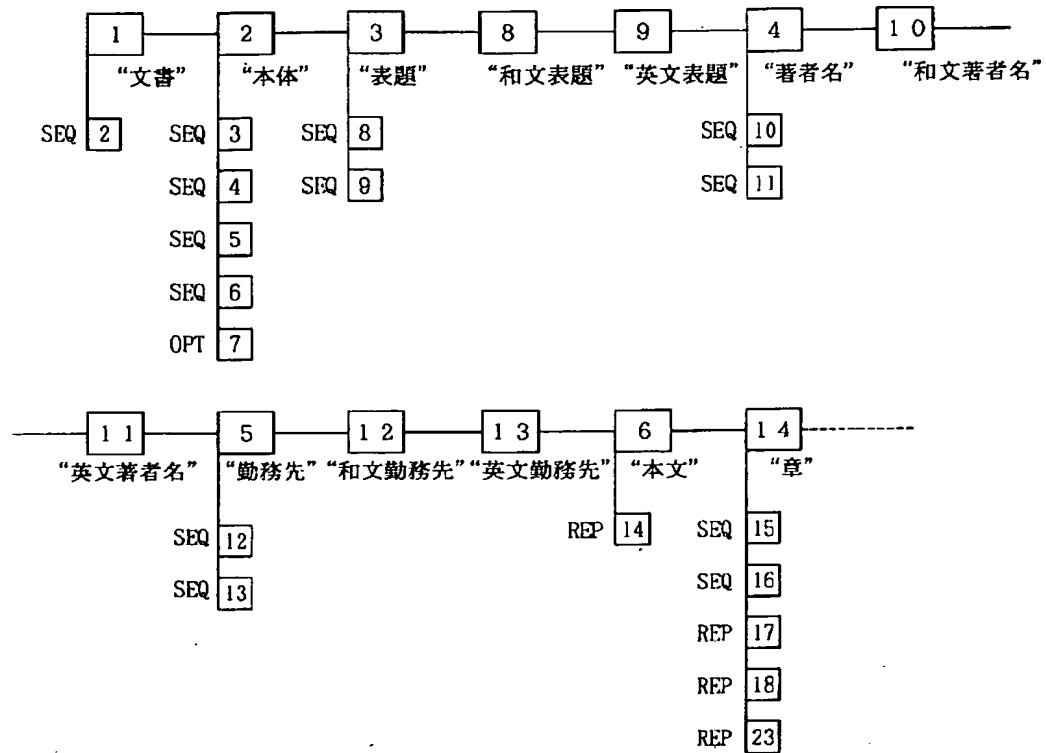
【図3】

## 共通論理構造



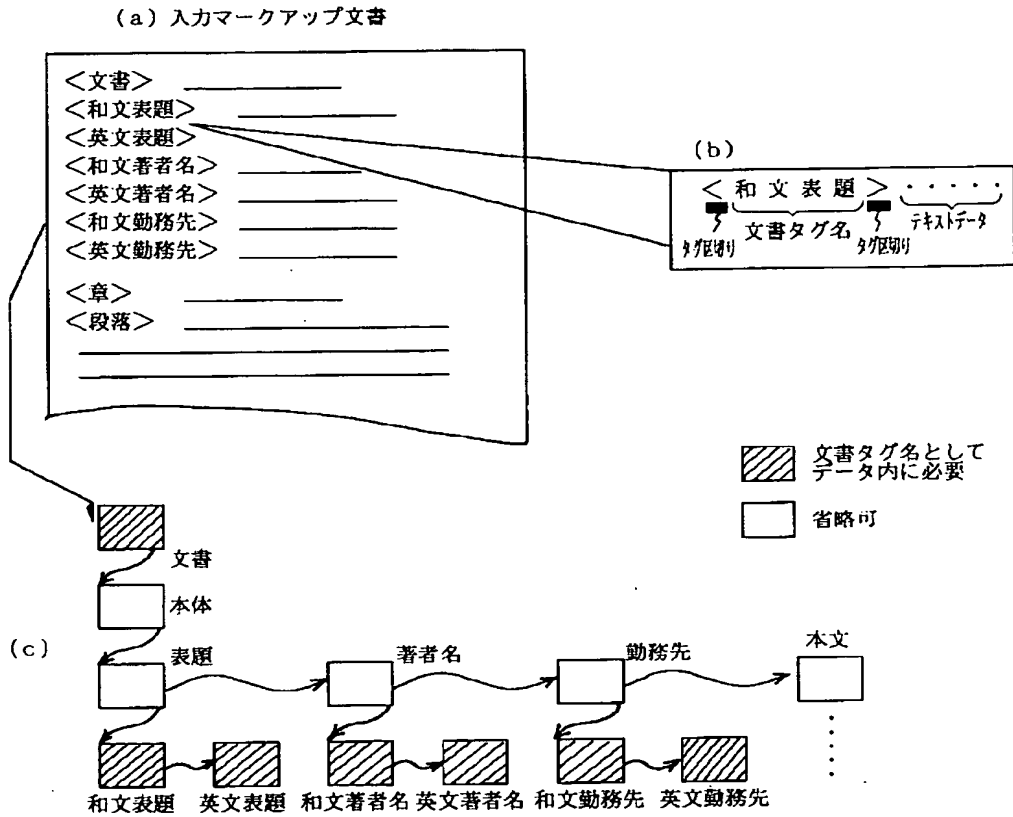
【図4】

## 展開後共通論理構造テーブルデータ

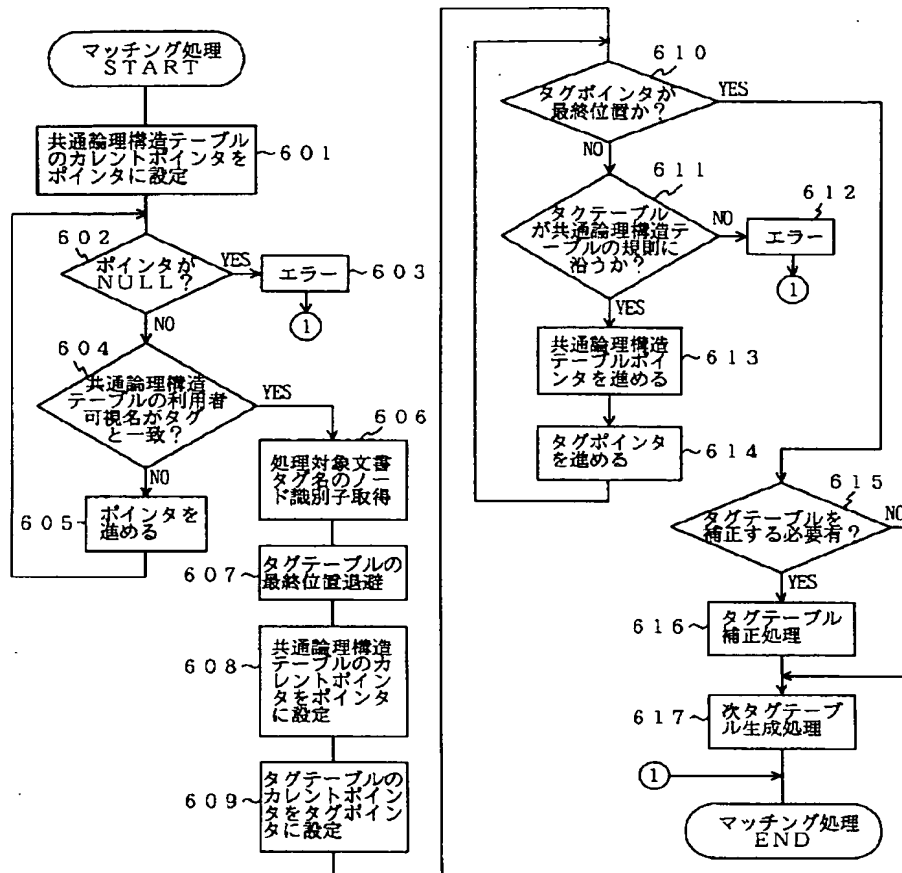




【図5】



【図6】



【図7】

## ●カレントポイント

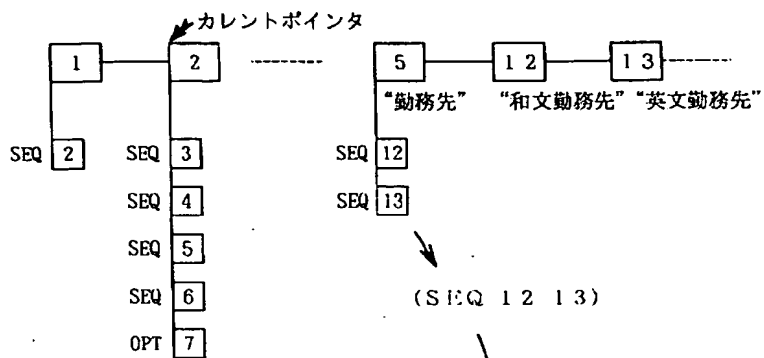


Diagram illustrating the relationships between author names and initials:

- Top Row:**
  - Box 1: 著者名 (Author Name)
  - Box 2: 勤務先 (Initials)
- Bottom Row (Left):**
  - Box 3: 和文著者名 (Chinese Author Name)
  - Box 4: 英文著者名 (English Author Name)
- Bottom Row (Right):**
  - Box 5: 和文勤務先 (Chinese Initials)
  - Box 6: 英文勤務先 (English Initials)

Arrows indicate the flow of information:

- From 著者名 to 勤務先.
- From 著者名 to 和文著者名.
- From 著者名 to 英文著者名.
- From 和文著者名 to 和文勤務先.
- From 英文著者名 to 英文勤務先.
- From 和文勤務先 to 勤務先.
- From 英文勤務先 to 勤務先.

フロントページの続き

(72)発明者 山崎 英二  
神奈川県横浜市中区尾上町6丁目81番地  
日立ソフトウェアエンジニアリング株式会  
社内